DOCUMENT RESUME

ED 099 425 95 TM 004 303

AUTHOR Larsen, Edwin P.
TITLE Opening Institutional Ledger Books--A Challenge to Educational Leadership: Suggestions for Talking to School-Community Groups About Testing and Test Results. TM Report No. 28.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ETS-TM-28
PUB DATE Dec 74
CONTRACT OEC-0-70-3797-519
NOTE 13p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Educational Testing; *Information Dissemination; *Manuals; School Districts; Schools; Scores; *Test Interpretation; *Test Results; Tests

ABSTRACT

Three key areas are outlined dealing with the development of public understanding of testing: (1) Why tests are administered in schools: needs assessment, instructional program evaluation, materials selection, reporting to public, documenting individual growth, diagnostic analysis and planning, and instructional grouping. (2) Types of tests used, featuring explanations of achievement tests, Scholastic Aptitude Tests, interest tests, specialized aptitude tests, and personality tests. (3) Interpretation of test norms, raw scores, grade equivalent scores, percentile ranks and stanines, I. Q. scores, and summarizing results (medians and quartiles). Methods used to chart test results of a school or district are discussed and suggestions made for the basic tools needed, the need for minimum use of numbers, and the facility of percentile ranks. Tables and charts for presenting statistical information are proposed, and suggestions include highlighting specific skills, comparing aptitude and achievement, and charting growth from grade to grade. Finally, in discussing results and school accountability, the following are proposed: assume leadership--an advocacy position in identifying discrepancies in pupil performance (needs), relate results to instructional efforts, discuss resource needs of the district and school, outline noninstructional problems the school and community must address, and approximate accountability. (RC)

ED 099425

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

TM REPORT 28 DECEMBER 1974

OPENING INSTITUTIONAL LEDGER BOOKS— A CHALLENGE TO EDUCATIONAL LEADERSHIP

Suggestions for Talking to School-Community Groups about Testing and Test Results

Edwin P. Larsen

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

INTRODUCTION

"Today we expect of the school that the majority of students really learn something—a novel, indeed unprecedented, demand."[1]

Businesses, particularly large ones, represent an idealized prototype of accountability to most of the American public. There is seldom a question of whether the stockholders and/or consumers have a right to information regarding the performance—or productivity—of a multimillion dollar enterprise. The popularity of this accountability concept is attested to by the wide support given to consumer advocacy groups and agencies.

Education is also big business, both nationally and at a local level. Education is our single largest national expense, and, because of this, its performance matters a great deal. Economists note that formal education in schools and colleges accounts for 12 percent of our gross national product. In my own community (Oakland, California), the public schools rank sixth among employers. In addition, schools have a near corner on the consumer market with 90 percent of the pupils attending public schools.

Regardless of the observations of Christopher Jencks,[2] our economy is now dominated by the "knowledge worker," so aptly described by Drucker. We have long passed the time when success within the academic institution matters little or not at all, when most people earned their livelihood as skilled craftsmen. Education will not guarantee success today, but without it success in our contemporary society will come with much more difficulty.

[1]Peter F. Drucker, *Rise of the Knowledge Worker*, Chicago, Encyclopedia Britannica, 1973.

[2]Christopher Jencks, *et al.*, *Inequality*, New York, Basic Books, 1972.

Little wonder then that both school stockholders (taxpayers) and school consumers (parents and students) ask school personnel (administrators and teachers) for information regarding the performance of the business (the educational program). These questions come in various forms:

Mother of fourth grader: "I've heard that one-half of the children in our school read below grade level. Is that true?"

Father (businessman): "Isn't it true that this new math program that we have in this school is a miserable failure? My son can't even add and subtract accurately."

Mother of junior high student: "Friends of ours are moving to this area. They have a tenth grader and want to know where to purchase a new home. What is the best senior high in this area?"

Community action group spokesman: "I believe that this district has failed to provide equal opportunities and resources for all youngsters. There are schools in this city where none of the children can read at all."

These are questions from consumers who do not have the facts about the academic performance of the students in the schools. Unfortunately, there are many school staff members who either do not have the information or are uncomfortable in discussing it with their constituencies. All too frequently, the professional establishment seeks to avoid this whole matter with the statement that "test data are not well understood and subject to misinterpretation by an unskilled public."

While there are many valid concerns regarding the potential misuse and misunderstanding of test scores, it

TM 004 633

This publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, represent official National Institute of Education position or policy.

is the writer's observation that test results are more frequently underused. I have heard students and parents express more concern about the nonuse and unavailability of test scores than about the misuses of results (labeling, tracking, and so forth).

Because certain legislative leaders in California felt school professionals were keeping test scores from public view, they took over the leadership in this area. It boiled down to "If the school people will not report to the public, we will." For several years, the State Department of Education has been required, under various state testing laws, to publish district-by-district and school-by-school results of state-required reading and math tests. There is less and less a question of whether test results will be disseminated to the public. The critical issue is "by whom?"

Public release of school-by-school test data by local school districts is not a common practice. In a number of school districts known to the writer, school-by-school averages are not even available to school principals and teachers. This was the case in my district seven years ago. In 1967, after two years of greater openness within the professional ranks, the first public school-by-school report was made to the Oakland Board of Education. More and more districts are taking this step. They are finding that there are many benefits to be realized in an atmosphere of openness and that the public and press are able to understand and use the results more wisely than was assumed possible.

The purposes of this paper are twofold: to present some background concepts regarding testing, scores, and statistics and to make some suggestions for communicating the information to other professionals, parents, students, and the public. It is hoped that some of the concepts and suggestions will be useful to school personnel intending to assume a leadership position in promoting better understanding and better uses of tests and test scores.

GROUP PRESENTATIONS OF TEST RESULTS

A Minicourse in Tests and Measurement

LESSON I: Development of Public Understanding of Tests and Testing

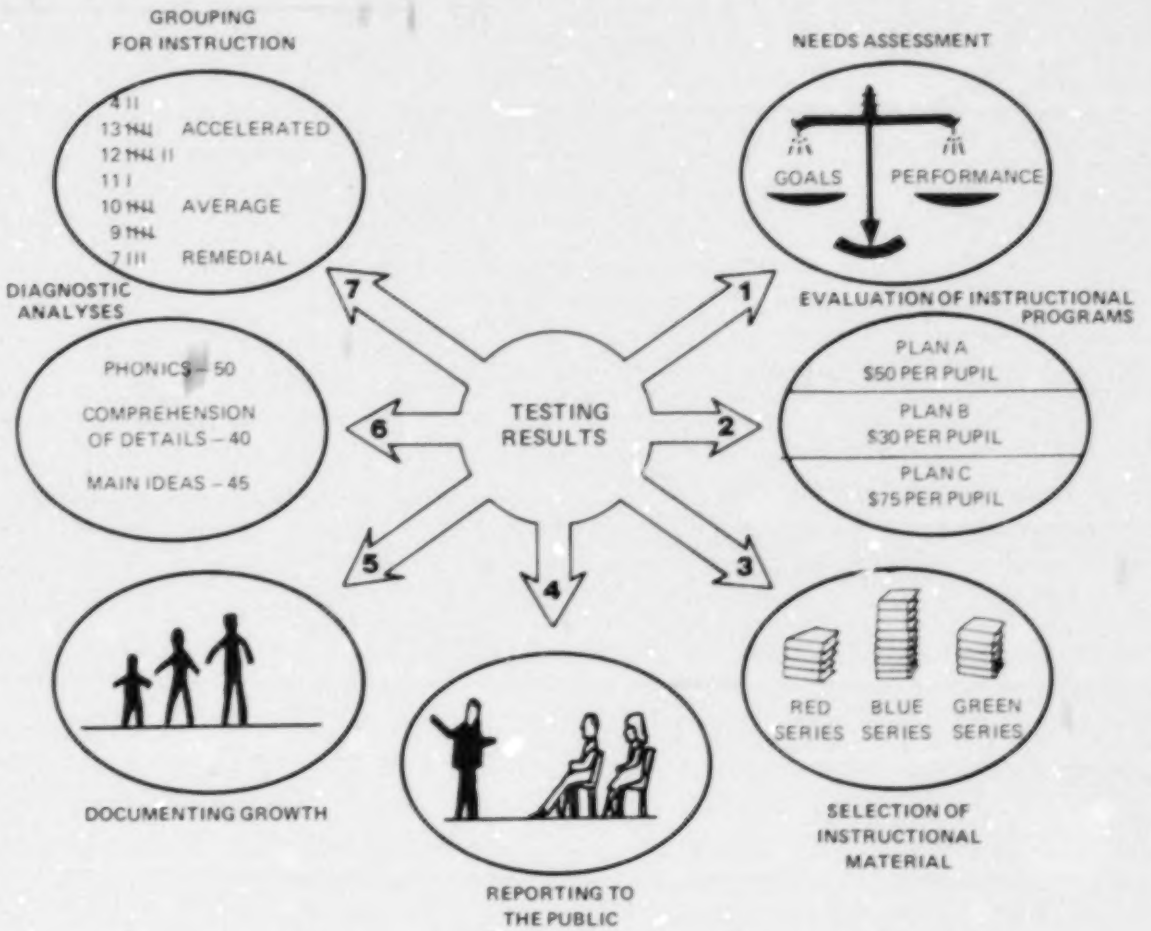
Following is an outline of the three key areas that should be addressed in making a group presentation of test data—whether it is to a professional, staff, or lay group. (Often one assumes that teachers or principals know much more about tests, statistics, and scores than they actually do.) The intent is to give the consumer an orientation or minicourse that will enable him to understand the results of a district or individual school. The depth into which one goes will vary with the group involved. A full presentation can take as long as an hour. Briefer presentations, however, are possible and sometimes necessary. Included in this report are several prototype figures, illustrations that could accompany a talk on testing. These figures are relatively easy to compose and, with modern audiovisual techniques, turn into overhead transparencies or slides. A few pieces of colored tape or a few watercolor pen strokes can liven up the transparencies considerably. It is also possible to cut out certain pages of this paper and put them through a transparency-making machine. Transparencies have proved useful in making a large number of public presentations of test results during the past few years. The figures will be referred to in the text that follows.

Background Area A: Why Are Tests Administered in Schools? There are a number of valid administrative and instructional reasons for administering group tests. You are off to a poor start if the only reasons you can give are that the state or district requires them. The only justification for administering tests or for anything else we do in schools is helping children. This does not always mean an individual pupil case study but may include using scores to assist in making a number of key planning decisions.

Listed below and illustrated in Figure 1 are some of the most common reasons for testing children. You will be well advised to present only those that apply to your school situation; platitudinous statements are sensed very quickly as being superficial. Each of these items can be tied concretely to showing how the information helps administrators, teachers, and parents to help children. Educational planning is an enormously complex task. Test scores provide only one type of input into the process. A test score, by itself, should never be the sole basis of a decision. However, test scores can contribute very useful information in the decision-making process.

1. *Needs Assessment.* Test scores are useful, objective indicators of pupil skill levels at a given point in time. The discrepancy between desired levels of pupil performance (group or individual) and actual performance constitutes an educational need or priority. When group or individual performance is at or above a level judged to be satisfactory, one has useful confirmation that the educational program, as it is functioning, is satisfactory. On the other hand, when group or individual per-

Figure 1: Why Are Tests Administered?



formance is below acceptable levels, one has a clear charge to make changes—more time, different materials, and/or different teaching strategies.

Test information can be useful to professionals and laymen in making important decisions regarding their plans for helping children. Examples can be given in almost any school or district where discrepancies have been found and program changes have been instituted or are planned.

Professionals and laymen sometimes express the opinion that test scores label children. This is a superficial observation since it is the professional or layman who does the labeling. Low performance should be discussed not hidden. It is only when the performance problems are discussed and appropriate actions taken that we can expect children to benefit or their performance to improve.

2. *Evaluation of Instructional Programs.* Educators are constantly searching for improved, more effective ways to help children. Pre- and posttest scores often provide critical information on the effectiveness of experimental and innovative programs. In many instances, schools are faced with finding programs that will be the most cost effective. With limited resources schools must find methods for maximizing the amount of learning resulting from each teaching hour. Federal and state agencies are demanding this type of accountability. Test scores can provide one type of information on the effectiveness of certain new programs. Examples such as the new

math programs or experiments in multiage grouping can be given in most school districts.

3. *Selection of Instructional Material.* Principals and teachers face the need to purchase instructional materials at difficulty levels appropriate for the pupils in their schools. Accelerated pupils will gain little from materials geared below their performance levels. Likewise, pupils performing at average and below-average levels need materials that are most appropriate to their current instructional levels. By examining the numbers of pupils achieving at various performance levels, administrators and teachers can make informed decisions regarding the purchase of supplies, textbooks, workbooks, and other materials best suited to the children.

4. *Reporting to the Public (Consumers).* Boards of education and citizens at large, as well as legislatures, continue to express interest in obtaining periodic reports of pupil performance. Many educational leaders have a keen interest in involving and informing the communities they serve. The purpose may be to build confidence in effective existing programs or to enlist support for needed improvements. Recent reports to boards of education as well as the very presentation being made to the group involved can be given as examples.

In some communities the increasing involvement of parents in planning and decision making makes it imperative that they be well informed. Parent recommendations founded on hearsay information are hardly sound bases for actions that will help children.

5. *Documenting the Growth of Individual Pupils.* Most parents and teachers have a great deal of information regarding the performance and growth of individual pupils. Many students are also aware of their status, owing to many feedback mechanisms (e.g., grades and their own observations). However, there are important instances where nonjudgmental (objective) information is needed by parents, students, and teachers.

Following are examples of questions that may be answered much more easily if test scores are available for individual pupils:

- "If we move to another part of the country, how will John fit into their program?"
- "I know Mary has good grades, but is she really college material?"
- "We feel your standards are lower than those of other schools. Just how well is Sue reading? Is she really at grade level?"

A Note of Caution. Group tests are not precision instruments for individuals; group averages are much more accurate than scores for one pupil. Therefore, individual scores should be viewed as reflecting approximate, or general, levels of performance. Some scores may be a little higher, others a little lower, than the student's true capacity.

There should be provisions at all schools for parents to have access to the scores of their children. Furthermore, there should be opportunities for pupils to receive feedback on performance. Methods may vary from school to school—it may be a teacher, counselor, or psychologist—but both parents and pupils have a right to this type of information, accompanied by wise interpretation and counsel. Some questions that are raised in an audience situation should be discussed in an individual conference. Your policies regarding interpretation of individual students' test scores should be explained.

6. *Diagnostic Analysis and Planning.* Most group achievement tests have been carefully designed to assess each of the most important skill areas with a sampling of only a few questions. There are not enough questions in each area to fully diagnose the strengths and weaknesses of individual children, but the tabulations of responses of a class or school group can often give diagnostic leads—i.e., the skill areas that have and have not been mastered. In some cases, it may mean reteaching materials already covered. In others, it may reveal key areas that have been overlooked.

7. *Grouping for Instruction.* Group achievement test scores can be useful in general planning for instructional groupings and in conjunction with the use of scores to select appropriate levels of textbooks. By examining the numbers of pupils performing in various ranges, one can develop plans for providing differentiated programs geared to the needs of the pupils.

Another Note of Caution. Group test scores should never be used as criteria for placing students into one group or another. It is well known that individual scores are subject to fluctuation (error). The practice of drawing a line, or cutting score, is an abrogation of professional responsibility. On more than one occasion, I have heard professionals explain to parents that it was not they who made a placement decision; it was a test score! An individual's score may be used for counseling or guidance; discussing how students who have previously entered a given algebra course with similar levels of entry skills (scores) have fared may help in making a decision. Or the score may be combined with other information (such as grades) and the preferences of the student in arriving at a pupil-oriented decision.

8. *Other Uses.* There are probably a number of other pupil-oriented uses made of test scores in your district. The foregoing represent only a few of the major areas that might be used to explain why tests are administered to pupils in the schools.

Background Area B: What Kinds of Tests Are Used? There is often an unnecessary air of mystery surrounding the tests and test scores. Unfortunately, some would like to keep it that way. It is very important to dispel any secretive aura that may exist by giving a few facts about the instruments that have been used. On the following pages are some approaches, or concepts, that may be used to explain the nature of the types of tests used most frequently in the public schools. These are shown in Figure 2.

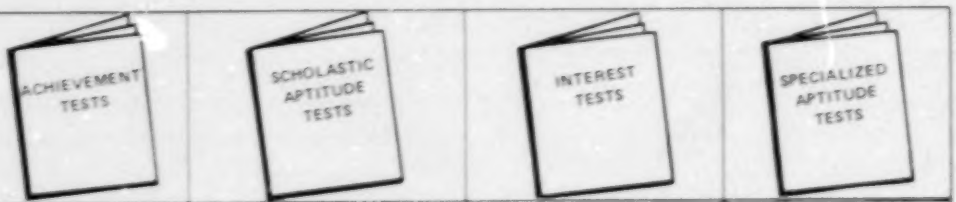
1. *Achievement Tests.* These tests are made up of a sampling of questions or problems on the types of materials covered and concepts taught in the curriculums common in schools throughout the country. They are not intended to cover only the specific curriculum of a given school or school district. Rather, the test is intended to assess how well a youngster can approach certain problem situations by applying the skills and concepts he has learned thus far in his academic career.

One should give specific examples of these questions or problems, items that are similar to those used in the tests. In my district, we have used orientation booklets that incorporate 3 to 6 sample questions from each subtest. These booklets were supplied by the publishers of the tests. There is seldom any puzzlement or objection to test content when parents or students see concrete examples of it. (Copies can be supplied for the CTBS and ITED upon request to the author.)

Overhead transparencies can be used if printed examples are not feasible. In some instances, a verbal explanation may be the best approach: "In a reading test, children are asked to read short paragraphs similar to those found in textbooks and to answer questions to show how well they have understood what has been read. Arithmetic or mathematics tests contain addition, subtraction, multiplication, and division problems. Also, they often contain questions to show how well students can solve word problems. A writing test usually instructs pupils to determine whether or not a sentence has proper punctuation, capital letters, and spelling. Pupils are also asked to select sentences that express an idea most clearly."

2. *Scholastic Aptitude Tests.* This is the second type of test most often administered in the schools. These tests are sometimes loosely referred to as intelligence or I.Q. tests. The test makers themselves prefer to consider these tests as measuring school aptitudes. The key assumption here is that a pupil's level of intellectual functioning, particularly in verbal and numerical skill areas, will give an indication, or prediction, of the rate of progress or success a pupil will have in future numerically and verbally oriented studies. The relationship between past

Figure 2: Types of Tests



	ACHIEVEMENT TESTS	SCHOLASTIC APTITUDE TESTS	INTEREST TESTS	SPECIALIZED APTITUDE TESTS
EXAMPLES	READING ARITHMETIC LANGUAGE	SCHOOL ABILITY READINESS I.Q.	VOCATIONAL INVENTORIES PREFERENCE TESTS	CLERICAL MECHANICAL SPACE RELATIONS MATHEMATICS
CONTENTS	TYPICAL PROBLEMS FROM BASIC SKILL AREAS. MEASURE ABILITY TO TRANSFER OR APPLY SKILLS LEARNED IN SCHOOLS	VARIETY OF VERBAL REASONING AND NUMERICAL REASONING PROBLEMS. MEASURE ABILITIES REQUIRED IN MOST SCHOOL COURSES	LISTS OF ACTIVITIES RELATED TO A VARIETY OF OCCUPATIONS. STUDENTS INDICATE PREFERENCES—E.G., OUTDOOR, PERSUASIVE, SCIENTIFIC, AND SO FORTH	REASONING AND SKILL PROBLEMS IN SPECIFIC SKILL AREAS JUDGED TO BE CRITICAL TO SUCCESS IN VARIOUS OCCUPATIONS
KINDS OF SCORES	GRADE EQUIVALENTS PERCENTILES (BASED ON NATIONAL SAMPLES OF STUDENTS)	PERCENTILES I.Q.s (BASED ON NATIONAL SAMPLES OF STUDENTS)	PERCENTILES OR OTHER RANKS (BASED ON SAMPLES OF STUDENTS OR SAMPLES OF PERSONS IN OCCUPATIONS)	PERCENTILES (BASED ON SAMPLES OF STUDENTS)

achievement and future achievement has been documented by hundreds of research studies.

Scholastic aptitude tests are more general in nature than achievement tests, but they must not be considered as direct measures of innate intelligence in the generic sense in which intelligence is generally discussed.

The home, school, and an individual's biological inheritance will have combined to affect how much verbal and numerical proficiency he has achieved and, hence, is able to demonstrate on a scholastic aptitude test. Intelligence is not directly measurable, nor are any other cognitive functions. Scholastic aptitude is merely judged by current achievement and/or skill levels.

Scholastic aptitude scores require very careful interpretation. A statement that a given group of pupils is performing "as well as can be expected," while accurate in a certain sense, is subject to much misinterpretation. Pupils may be performing on reading or arithmetic tests as well as can be expected, given all the antecedent conditions that contributed to their present verbal and numerical skill levels. This does not say that "this is all the pupils are capable of."

There is, at best, a vague distinction between achievement and scholastic aptitude tests. Virtually identical questions may arise concerning both.

In this writer's view, very cautious use should be made of scholastic aptitude or I.Q. scores. Given a certain level of general verbal and numerical abilities (a scholastic aptitude score), is a student able to apply his skills to problem situations (achievement test items) as well as other pupils who have approximately the same verbal and numerical aptitude scores? If a person does not feel comfortable in dealing with the interpretation of scholastic aptitude scores, he would be advised to omit them and deal with achievement scores only.

It is suggested that samples of scholastic aptitude tests be shown to an audience so that they will fully understand the bases of these scores.

3. *Interest Tests.* Some school districts administer vocational interest tests as part of programs designed to promote vocational readiness and awareness of career options. Generally, students are asked to express preferences for various kinds of activities and experiences. The vocational interest inventory provides scores that summarize interests into types (persuasive, social welfare, artistic, and the like) or show how a person's responses compare to those of persons who are successful and satisfied in various occupational categories. These tests are not intended to tell a student what vocation to pursue. Rather, they yield some facts that can be used along with student achievement, family values, and so forth in making vocational choices.

4. *Specialized Aptitude Tests.* Some tests are designed to measure very specialized aptitudes, such as mechanical, clerical, spatial relationship, mathematical, and abstract reasoning skills. A report on how well a student has achieved in these areas may also help him or her make decisions about entering fields such as engineering or business.

5. *Personality Tests.* Instruments that attempt to measure a person's innermost psychological makeup and attitudes are seldom administered in public schools. Parental permission is usually required, and such tests are administered only by trained psychologists. In many states, such testing is controlled by law, but some parents wonder about such "invasions of privacy," and it may be appropriate to explain local practices concerning this area of testing.

Background Area C: How Do I Interpret Different Kinds of Scores and Statistics? Some basic definitions are in order before moving on to an examination of score averages for a school and/or a district.

1. *Test Norms (Making Comparisons).* The most common types of standardized test scores are those that compare a pupil's performance with the performance of children in a national norms group. The publisher of a test selects a sample of children from all parts of the nation in order to estimate the average or typical level of performance of children at a given grade level. This sampling procedure is similar to the one used in opinion polls. The information obtained from this sample group becomes the publisher's norms, and we can determine how a child or group of children compares with this sample group.

In California at certain grade levels, a pupil's scores can also be compared with scores of other children in the same grade throughout the state. The state score patterns and averages based on testing all children in the state at selected grade levels are used as state norms. We can tell, for example, how a sixth grader ranks in reading ability with other sixth graders throughout the state.

The score averages of the city provide yet another comparison in evaluating local school results. The question here is "How well does a pupil or school rank within the city?"

One can see that it is important to keep in mind the group with which a pupil's performance is being compared.

2. *A Raw Score (Number Correct) Does Not Mean Much.* Two types of scores are generally used to evaluate test performance. It is probably better to select one or the other for purposes of presentation. Some ideas in the following section may be useful in explaining these derived or converted scores to clients.

- *Grade Equivalent Scores.* Grade equivalent scores, sometimes called grade placement scores, are often used because the number of correct answers on a test has little meaning in itself. A score of 20 on one test may show first grade performance. On another test the same score may show high school level skill development. Much depends on the test's level of difficulty. In order to develop grade equivalent scores, publishers administer the tests to children in different grade levels during a given school month.

The average number correct, or raw score, earned by first graders in the ninth month of school is assigned a grade equivalent value of 1.9. The average raw score for third graders in the ninth month would be assigned a grade equivalent value of 3.9. We can, therefore, compare a pupil's raw score with the scores of other children. One will note in Figure 3 scores such as 1.3,5.6, and so on. The number 1.3 refers to the score earned by an average first grader at the third month of the school year, and 5.6 means a score level typical at the sixth month of grade 5. To answer the question "What is an average score?", one must indicate the grade and month of the school year in which the test was administered.

Figure 3. Interpretation of Grade Equivalent Scores

NUMBER CORRECT	GRADE EQUIVALENT	INTERPRETATION
20	4.3	
19	4.1	(SCORE EARNED BY AVERAGE
18	3.9	(THIRD GRADER IN NINTH MONTH
17	3.7	(OF GRADE 3.
16	3.5	
15	3.3	
14	3.1	(SCORE EARNED BY AVERAGE
13	2.9	(SECOND GRADER IN NINTH MONTH OF
12	2.7	(GRADE 2.
11	2.5	
10	2.3	
9	2.1	(SCORE EARNED BY AVERAGE
8	1.9	(FIRST GRADER IN NINTH MONTH OF
7	1.8	(GRADE 1.
6	1.7	
5	1.6	
4	1.5	
3	1.4	
2	1.3	
1	1.2	
0	1.1	

- *Percentile Ranks and Stanines.* Percentile ranks are also useful in interpreting test performance. Percentile rank is not the same as percent correct, which is often seen on teacher-made tests such as in spelling or arithmetic. A percentile rank indicates the percentage of students within a reference or norms group whose scores fell at or below a given score.

Students differ widely in their physical and mental performance. If we were to take 100 randomly selected youngsters from all parts of the nation and ask them to run a race, we would soon find differences in running skills. The same is, of course, true with reading, arithmetic, and other skill areas.

The 50th percentile is considered to be "right on the grade level" or right at the average of the norms group. A pupil who is at the 50th percentile in grade 5 is at the average grade level for fifth graders. Similarly, a pupil with a 50th percentile rank in grade 8 is at the average grade level for eighth graders but is achieving at a higher level than the average fifth grade youngster.

In evaluating the test performance of an individual student, one must keep in mind that a given score is only one sample of the student's ability to perform in a given subject area. An individual student's scores would generally vary considerably if he were tested several times with tests of equal difficulty levels and covering essentially the same skill areas. Variations in scores are generally due to variations: (1) in the way a student approaches the test (his motivation or his physical condition at the time of testing); (2) in the way the test was administered (how well directions were explained or the general conditions under which the student was working); and (3) in the tests themselves (tests cannot include items measuring every skill taught or learned in a given area; rather, each test contains what is hoped to be a balanced sampling of items. One test may sample more of what the student knows than another test may sample).

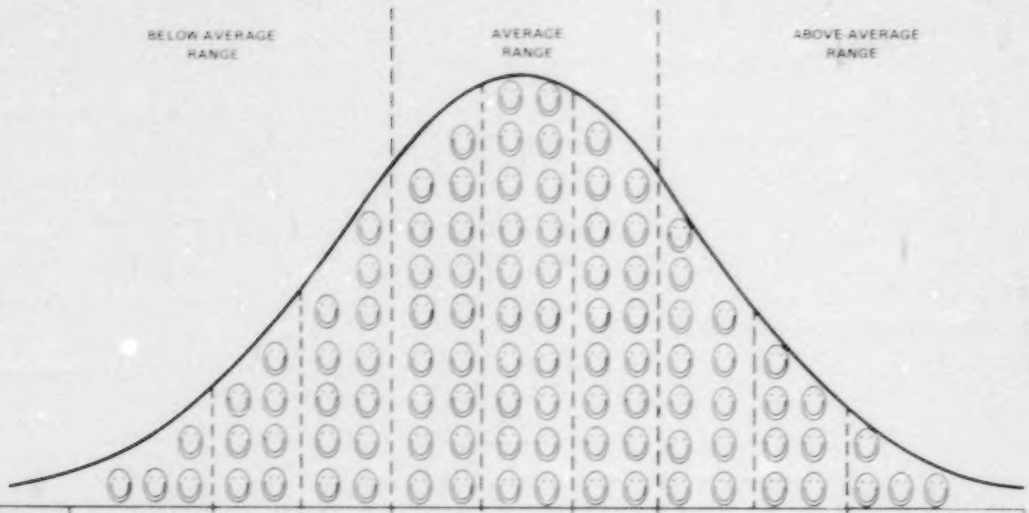
Therefore, a percentile rank between 25 and 75 is usually considered to be within an average range. Variations of several points in individual student scores should not be misinterpreted.

Figure 4 may be used to give a graphic illustration of how a typical group of pupils would score on a 20-item reading test. The 100 students who are illustrated represent the distribution or spread of scores within a hypothetical norms group. Actually, publishers utilize the scores of several thousand youngsters in deriving percentile ranks.

One will note that the percentile rank range has been divided into nine levels called stanines. In evaluating differences in percentile ranks falling within a given stanine or stanines, a good rule of thumb is to discount differences of one stanine level or less. Differences of more than one stanine level are probably significant variations in performance. The following examples may serve to illustrate this point:

> John received a percentile rank of 28 on the reading test, which is within the stanine 4 level, and a percentile rank of 53 on the math test, which is within the stanine 5 level. His performance on the reading and math tests should be considered to be comparable. We have little proof that score differences of this magnitude reflect true differences in skill levels; they may have resulted from the sources of error noted previously.

Figure 4: Percentile Ranks



TEST RAW SCORE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PERCENTILE RANKS	1	2	4	6	9	14	20	26	34	44	54	64	72	80	86	90	94	97		99+	
STANINES	1			2		3		4		5		6		7		8		9			
PERCENTS IN STANINES	4			7		12		17		20		17		12		7		4			

Mary received a percentile rank of 8 on the math test (stanine 2 level) and a percentile rank of 26 on the reading test (stanine 4 level). Mary is probably significantly more skillful in reading than in math since her respective percentile ranks differed by more than one stanine level.

- *I.Q. Scores.* For many years the term I.Q. has been a common word in our society. Unfortunately, the meaning of this term is not well understood. An I.Q. score is just another type of score used to indicate where a person's performance ranks in comparison with the performance of individuals in the publisher's norms groups.

  I.Q. tests are not intended to be measures of native intellectual ability, as is sometimes assumed. Rather, the ones generally administered in schools sample student skills in verbal and numerical reasoning areas. A better example of the group "I.Q." tests used in schools is probably a scholastic aptitude test.

  An I.Q. score of 100 is used by test makers to represent the score level of an average person in a given age group. Approximately 68 percent of a given age group receive I.Q. scores between 84 and 116 on most modern scholastic aptitude tests. The I.Q. score system is based on statistical assumptions that are not commonly understood. Therefore, it is probably easier to grasp where a person ranks on a scholastic aptitude or I.Q. test when scores are expressed as percentile ranks rather than as I.Q.s.

3. *Summarizing Results (Medians and Quartiles).* As we have seen, there is almost always quite a spread in pupil scores. Some pupils may get nearly all questions correct, others very few. The distribution of scores on a reading test, as shown on following page, will illustrate this fact. To summarize this array of scores, median and quartile statistics are usually used.

- Q1, or the first quartile, indicates the point at or below which one quarter or 25 percent of the pupils fell. Three-fourths of the pupils scored higher.

- The median, or second quartile, indicates the mid-point in the score range. One half of the students are above and one half below this point. This is important to keep in mind. A median obviously does not reflect the scores of all children.

- Q3, or the third quartile, indicates the point at or below which three quarters or 75 percent of the children have scored. One-fourth have scored higher.

Figure 5 can be used to show a sample tally of scores for a group of 96 pupils. One can also show the quartiles and the publisher's grade equivalents and percentile ranks for each quartile.

Figure 5. Distribution of Reading Test Scores

Publisher's Information		Data from School X			
Publisher's Grade Equivalent GE Scores	Publisher's Percentile Ranks	Number Correct	Number of Students Tested	cum f	
7.2	99	20	I	96	
6.8	99	19	II	95	
6.4	98	18	I	93	
5.9	97	17	III	92	
5.3	94	16	II	89	Q3—(¾ or 75% were at or below a score of 13, or GE 4.4)
5.0	90	15	TNL I	87	
4.7	86	14	TNL	81	
4.4	80	13	TNL III	76	
4.2	72	12	TNL TNL II	68	
4.0	64	11	TNL TNL I	56	Median or Q2—(½ or 50% were above or below a score of 11, or GE 4.0)
3.8	54	10	TNL III	45	
3.6	44	9	TNL TNL	37	
3.2	34	8	TNL	27	
2.9	26	7	TNL II	22	Q1—(¼ or 25% were at or below a score of 8, or GE 3.2)
2.7	20	6	TNL	15	
2.5	14	5	III	10	
2.1	9	4	III	7	
1.7	6	3	I	4	
1.2	4	2	II	3	
1.0	2	1	0	1	
1.0	1	0	I	1	
		Total Students	96		

LESSON II: Charting Your Results

The specific method used to chart, or present, the test results of a school or district will depend on the points one wishes to highlight. The illustrative charts described in this section show only some of the comparisons or analyses that can be made. In most instances, the audience is assumed to be an individual school-community group.

Background Considerations: Preplanning. There are a number of things to consider in preplanning:

1. *Basic Tools.* You will need a frequency distribution of the district and/or school results you intend to present. If one is not available through a district or commercial scoring service, a hand tally will need to be done since virtually any summary presentation focuses on the distribution of scores. You will also need the quartile points (Q1, Q2/median, and Q3). Raw scores are the best starting point for summarizing the results for any test, but to become meaningful they must be converted to percentiles, grade equivalents, or another norm-referenced score system.

The other basic item you will need is a test manual and/or handbook that contains the norms tables for the test for the time of year at which the tests were administered.

2. *Minimize the Use of Numbers.* Use graphic representations of percentile ranks, grade equivalents, or I.Q. score ranges whenever possible. Converted scores are abstractions, and graphic presentations generally give most of the information staff and parents want.

Quartiles are commonly used and easily understood summary statistics. Bar graphs showing the range between the first and third quartiles as well as the median are simple and direct methods of summarizing results. The use of medians without the presentation of first- and third-quartile data tends to give a very narrow presentation of the data. It is critical that parents and staff recognize the variability in student performance not only at the local level but also within the norms group. By focusing on the performance of the middle 50 percent of the population, one is able to moderate the "everyone should be above grade level" concept that dominates the thinking of many individuals. Everyone in a group just cannot be above the middle level performance of the group.

Several of the sample charts discussed in the next section utilize simple bars to show the interquartile range. This is the range between Q1 and Q3. The median, or Q2, is shown by a horizontal line across the bar.

3. *Percentile Ranks Are Easier.* To chart results for any test at any grade level or for multiple grade levels, a standard chart format can be used. Furthermore, the norms group quartiles are always located at constant reference points (Q1 = 25th percentile; Q2 = 50th percentile; Q3 = 75th percentile). If an audience can grasp this standard frame of reference, it will be much easier to present data for different tests and grade levels. In addition, percentile ranks are a valid base for comparison of results for different tests or subtests.

Grade equivalents may be more appropriate for your purposes, but the development of bar graphs is much more difficult because of the [illegible] range of scores (e.g., a first grade test may yield scores ranging from 1.0 to 5.0; a junior high school level test may yield scores from 2.0 to 13.0).

Sample Tables and Charts. The following tables and charts should give some idea of ways to present statistical information.

1. *Results for a Single Test—Comparing Norms Group, District, and School Results.* Figure 6 shows a sample of a combined statistical (numeric) and graphic presentation. It will be noted that results are shown for a single test at a single grade level. Since percentile ranks are used, no elaborate scaling is needed for the chart. The key here is the approximate relationship of the local interquartile range to the norms group. Only enough numbers are given to define the quartiles. One can see at a glance that, while the district median is two percentile points below the 50th percentile rank of the norms group, that is not the performance level of the entire

Figure 6. Percentile Ranks

	NORMS GROUP	DISTRICT	SCHOOL
Q_3	75	73	78
Q_2 (MEDIAN)	50	48	53
Q_1	25	22	20

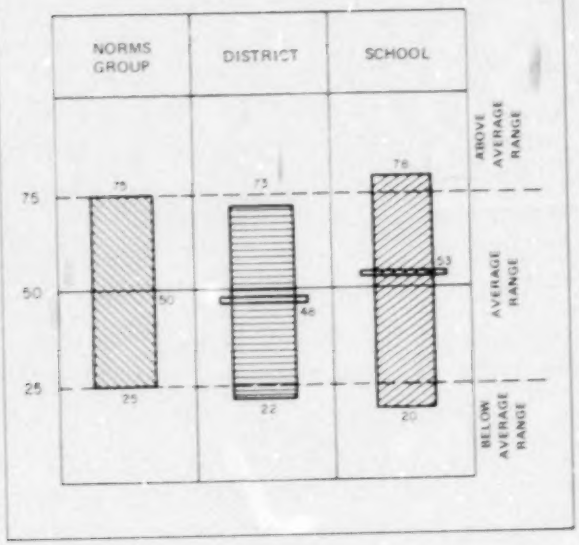
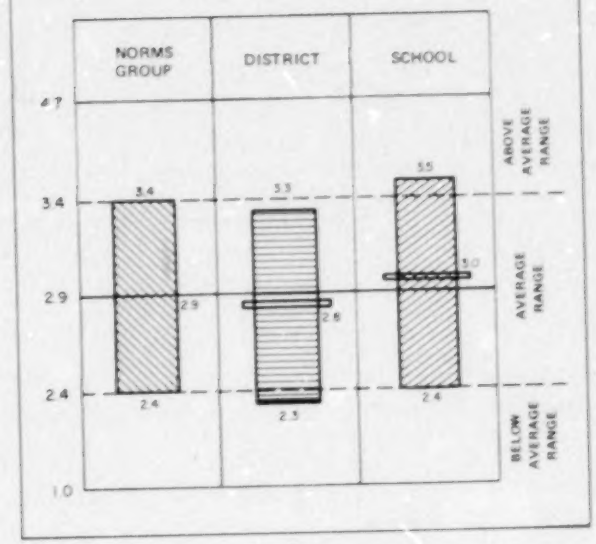


Figure 7. Grade Equivalents

	NORMS GROUP	DISTRICT	SCHOOL
Q_3	3.4	3.3	3.5
Q_2 (MEDIAN)	2.9	2.8	3.0
Q_1	2.4	2.3	2.4



school population. In fact, nearly one-fourth of the district population is scoring in the top quartile range of the norms group.

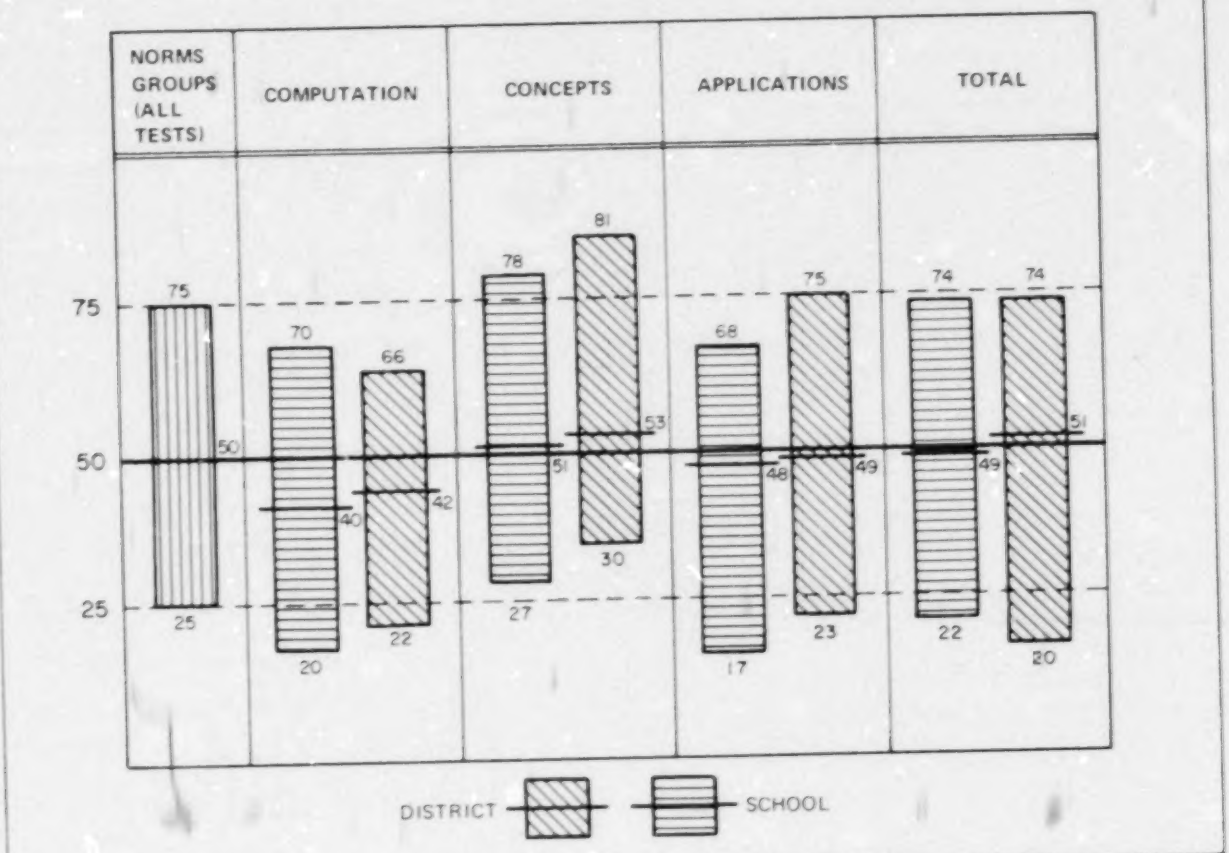
The same hypothetical data are presented in grade equivalent form in Figure 7. To determine the grade equivalent values for the quartiles (Q1, Q2, and Q3) of the norms group, use the publisher's test manual. Find the raw score equivalent for the 25th, 50th, and 75th percentiles, respectively. Then convert these raw score values into grade equivalents. These grade equivalent values will determine the upper and lower limits of the norms group bar and the median.

2. *Highlighting Specific Skills—Comparing Performance on Subtests.* Figure 8 illustrates a data presentation intended to analyze the comparative performance of students in different skill areas. In this sample, subtests from an arithmetic test are used. Results of subtests such as reading, spelling, and language usage could also be compared in this manner.

A school staff or board of education or parent group could see quickly the relative strengths and weaknesses of the groups tested. For example, students in both the district and school performed slightly higher than the norms group in Concepts, somewhat below in Computation, and slightly below in Applications skills. The overall performance of the district was only slightly below that of the norms group; the school's performance approximated that of the norms group at median and third quartile points but showed proportionately more youngsters in the lowest percentile rank levels than the norms group. These data might be of particular interest if a new mathematics curriculum were under study. They also might help a mother understand her child's individual test profile.

3. *Comparing Aptitude and Achievement.* In Figure 9, selected percentile rank ranges, called stanines, have been used to present the comparative performance of the publisher's norms groups and that of a school's student population on an aptitude test and an achievement battery. By definition there are set percentages of students scoring within each stanine level in the publisher's norms groups. Some districts utilize stanines as

Figure 8: Subskill Scores



the basic standard score for reporting individual and/or group scores, so this might be an appropriate format in some instances.

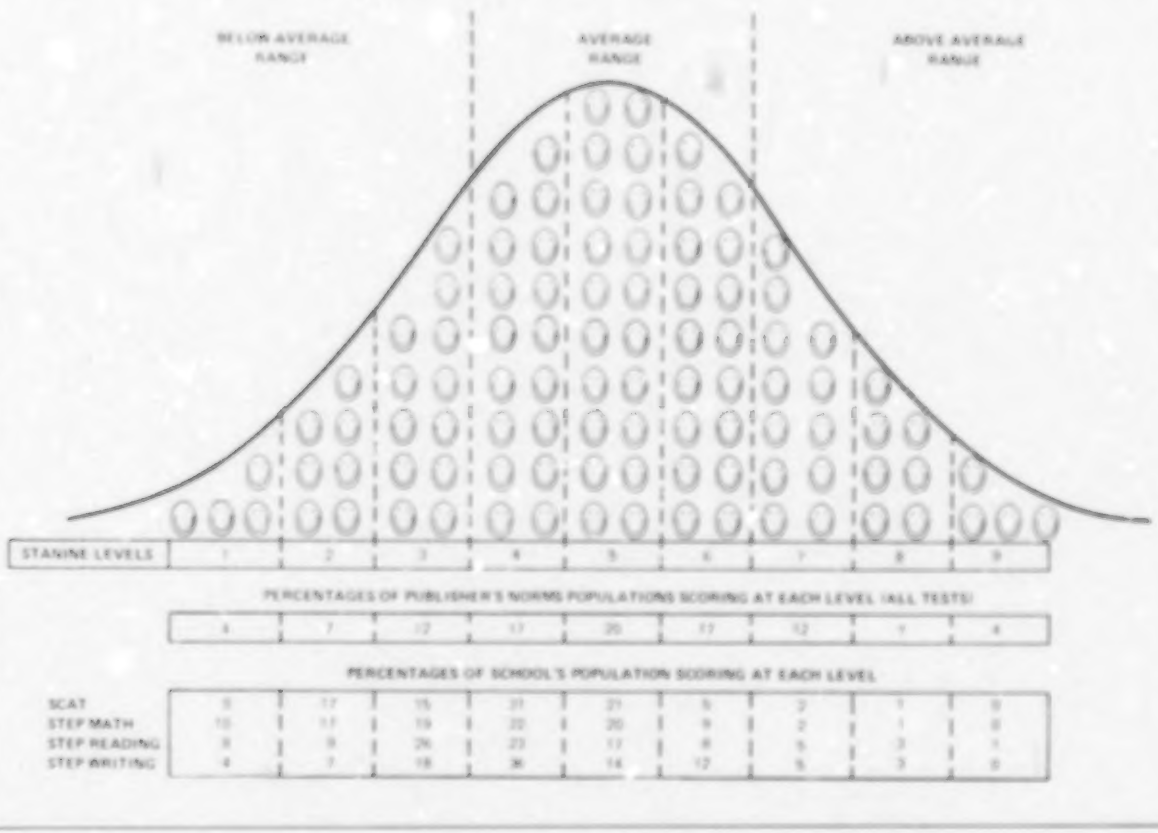
Key points that can be illustrated by Figure 9 are:

- Students show varying levels of scholastic aptitude and achievement in the publisher's norms population as well as at a given school.
- Even though average or median performance may be above, at, or below that of the norms group, there are diverse student needs in the district.
- Students' skill levels (achievement scores) are slightly lower than general scholastic aptitude scores would suggest they are currently capable of.
- It is relatively simple to plot the distribution of this school's performance in curve form during a presentation using a transparency marker. The curve would show graphically the levels at which the school's score patterns differ from those of the norms population.

At the sample school, median performance for all tests would be in the fourth stanine range. For each test presented we reach 50 percent of the school's population after we get out of stanine 3 and before we get to stanine 5. For example, consider SCAT. We add the 8, 17, and 15 percent in stanines 1, 2, and 3 respectively, and we have 40 percent. Since 31 percent of the school's population falls in stanine 4, we reach the median somewhere in this stanine range. However, there are youngsters whose achievement levels vary substantially from the median level. A small proportion are performing at substantially above average levels. For example, 9 percent of this school's population are in stanines 7, 8, and 9 on STEP Reading. A larger proportion of the school's youngsters are performing at below average levels: On STEP Math, for instance, 46 percent of the youngsters are in stanines 1, 2, and 3.

4. *Charting Growth From Grade to Grade (Norms Group and District).* There are occasions when it is desirable to show data on student progress from grade to grade. Unfortunately, few school districts have the resources necessary to compile student-by-student data over many years.

Figure 9 Aptitude and Achievement


BELOW AVERAGE RANGE
AVERAGE RANGE
ABOVE AVERAGE RANGE
STANINE LEVELS 1 2 3 4 5 6 7 8 9
PERCENTAGES OF PUBLISHER'S NORMS POPULATIONS SCORING AT EACH LEVEL (ALL TESTS)
4 7 12 17 20 17 12 7 4
PERCENTAGES OF SCHOOL'S POPULATION SCORING AT EACH LEVEL
SCAT 5 12 15 21 21 5 2 1 0
STEP MATH 10 11 19 22 20 9 2 1 0
STEP READING 5 9 26 23 11 8 5 3 1
STEP WRITING 4 7 18 30 14 12 5 3 0

Cross-sectional data for various grade levels tested in the same year or cross-sectional data for a given grade cohort tested year after year can be used to approximate such growth studies. As a matter of fact, these types of data are probably as valid as any, since the norms data for developing grade equivalents, percentiles, and so forth are cross-sectionally derived.

Figure 10 suggests a method for charting the interquartile ranges of data described above. Such a presentation may be useful in showing a variety of trends, such as:

- grades during which students are or are not showing desired amounts of growth;
- students who may be below grade level are, nevertheless, making substantial growth gains.

LESSON III: Discussing Your Results? How Accountable Are Schools for Test Scores

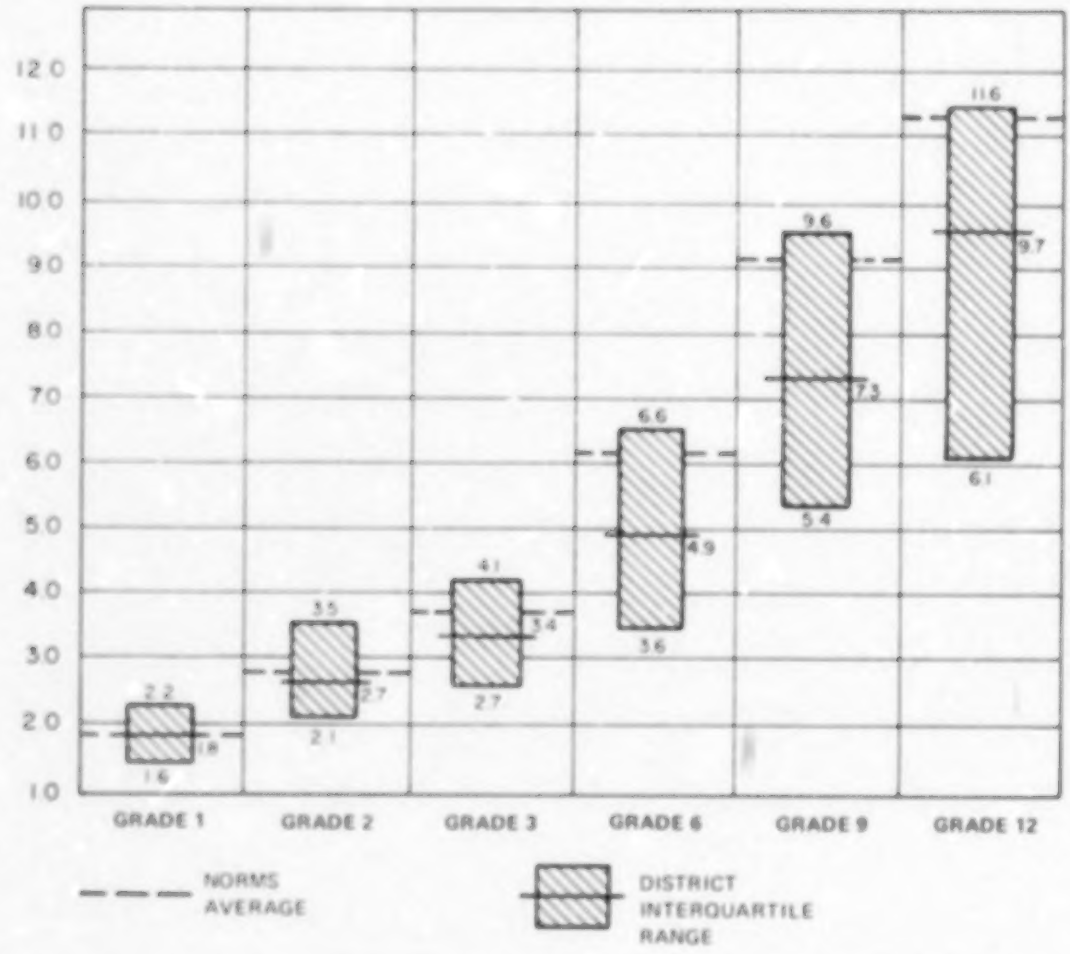
How accountable the schools are is obviously an unanswerable question—at least in simple terms. However, the following suggestions are presented for those who may find themselves in the accountable spotlight.

1. *Do Not Explain the Test Scores Away.* If you emphasize that the tests were not administered properly or that you are convinced that they are totally irrelevant, you will immediately confirm the suspicion of many consumers.

2. *Assume Leadership—An Advocacy Position in Identifying Discrepancies in Pupil Performance Needs.* This is not to say that one can promise instant remediation of performance problems. However, I have seldom seen a person attacked by a consumer group if he is showing an intention to act in response to these needs.

3. *Relate Results to Instructional Efforts.* The results mean little in isolation. Even though strong nonschool factors may be prominent influences, one can hardly take the position that the school program is not responsible, at least in part, for student performance. Known program strengths and weaknesses, highlighted by the results, should be discussed.

Figure 10. Grade to Grade Growth



4. *Discuss Resource Needs of the District and/or School.* Needs for supplies, facilities, and other fiscally constrained items should be noted. On the other hand, one may point to benefits derived from resources made available in the past.

5. *Outline Noninstructional Problems the School and Community Must Address.* Some of these might include:

- Absenteeism—including excused and unexcused absences. These often constitute significant problems in a number of schools.
- Physical well-being of pupils—including problems associated with nutrition, adequacy of clothing, and treatment of major medical problems.
- Environmental complement to school instruction—including the paramount importance of parent interest in the child's school activities, encouragement of reading at home, importance of discussing with and explaining to children (building language competence).
- Pupil interest and motivation.
- Pupil mobility.

6. *Approximate Accountability.* Describe the school's efforts in attempting to meet some of the problems listed above. Communicating the feeling that the schools are also recognizing problems related to instruction (e.g., "This school needs to strengthen its arithmetic computation skills.") as well as problems attributed to the population they serve is essential to building confidence in the schools. Judiciously acknowledging weaknesses where they exist and indicating that efforts are being made to correct them are the strongest positions that can be taken. This may be the closest the schools can come to being accountable. I believe that it is about all that our consumers expect of us.